

日 本 国 特 許 庁  
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日                      2 0 0 2 年 1 1 月 2 1 日  
Date of Application:

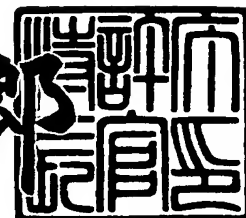
出 願 番 号                      特 願 2 0 0 2 - 3 3 8 4 9 7  
Application Number:  
[ST. 10/C]:                      [ J P 2 0 0 2 - 3 3 8 4 9 7 ]

出 願 人                      インターナショナル・ビジネス・マシーンズ・コーポレーシ  
Applicant(s):                      ョン

2 0 0 3 年    7 月 1 0 日

特許庁長官  
Commissioner,  
Japan Patent Office

太田信一郎



【書類名】 特許願

【整理番号】 JP9020204

【提出日】 平成14年11月21日

【あて先】 特許庁長官 殿

【国際特許分類】 G11B 20/10  
G06F 3/06  
G06F 5/06

【発明者】

【住所又は居所】 神奈川県藤沢市桐原町 1 番地 日本アイ・ビー・エム株式会社 藤沢事業所内

【氏名】 山田 秀二

【発明者】

【住所又は居所】 神奈川県藤沢市桐原町 1 番地 日本アイ・ビー・エム株式会社 藤沢事業所内

【氏名】 斉藤 博史

【発明者】

【住所又は居所】 神奈川県藤沢市桐原町 1 番地 日本アイ・ビー・エム株式会社 藤沢事業所内

【氏名】 斎藤 高裕

【発明者】

【住所又は居所】 神奈川県藤沢市桐原町 1 番地 日本アイ・ビー・エム株式会社 藤沢事業所内

【氏名】 福久 良司

【特許出願人】

【識別番号】 390009531

【氏名又は名称】 インターナショナル・ビジネス・マシーンズ・コーポレーション

## 【代理人】

【識別番号】 100086243

【弁理士】

【氏名又は名称】 坂口 博

## 【代理人】

【識別番号】 100091568

【弁理士】

【氏名又は名称】 市位 嘉宏

## 【代理人】

【識別番号】 100108501

【弁理士】

【氏名又は名称】 上野 剛史

## 【復代理人】

【識別番号】 100104880

【弁理士】

【氏名又は名称】 古部 次郎

## 【選任した復代理人】

【識別番号】 100118201

【弁理士】

【氏名又は名称】 千田 武

## 【手数料の表示】

【予納台帳番号】 081504

【納付金額】 21,000円

## 【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9706050

【包括委任状番号】 9704733

【包括委任状番号】 0207860

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 データ記憶装置及びバッファメモリの管理方法

【特許請求の範囲】

【請求項 1】 磁氣的にデータを記録する磁気ディスクと、

前記磁気ディスクに対して読み書きされるデータを当該磁気ディスクへの処理に先立って一時的に保持するバッファメモリとを備え、

前記バッファメモリは、記憶領域を、外部から受け付けた前記磁気ディスクへの書き込み要求を格納する第 1 のリングバッファと、前記磁気ディスクから読み出したデータを格納する第 2 のリングバッファとに分け、当該第 1 のリングバッファ及び当該第 2 のリングバッファを分ける仕切りの位置を可変とすることにより、当該第 1 のリングバッファ及び当該第 2 のリングバッファのバッファサイズを可変としたことを特徴とするデータ記憶装置。

【請求項 2】 前記仕切りは、前記第 1 のリングバッファまたは前記第 2 のリングバッファのボトムページであることを特徴とする請求項 1 に記載のデータ記憶装置。

【請求項 3】 前記バッファメモリは、前記磁気ディスクへの外部からのアクセスパターンに応じて前記仕切りを移動することを特徴とする請求項 1 に記載のデータ記憶装置。

【請求項 4】 データを記録する記録媒体と、

前記記録媒体に対する外部からの書き込み要求を当該記録媒体への処理に先立って一時的に保持する第 1 のリングバッファ及び前記記録媒体に対する外部からの読み出し要求に応じて当該記録媒体から読み出されたデータを保持する第 2 のリングバッファを持つバッファ手段と、

前記記録媒体への外部からのアクセスパターンに応じて、前記バッファ手段における前記第 1 のリングバッファ及び前記第 2 のリングバッファのバッファサイズを変更するバッファ制御手段と

を備えることを特徴とするデータ記憶装置。

【請求項 5】 前記バッファ手段は、バッファ領域を前記第 1 のリングバッファと前記第 2 のリングバッファとに分割した R A M (Random Access Memory)

であることを特徴とする請求項 4 に記載のデータ記憶装置。

【請求項 6】 前記バッファ制御手段は、前記 RAM のバッファ領域を分割する前記第 1 のリングバッファのボトムページまたは前記第 2 のリングバッファのボトムページの位置を変更することにより、前記第 1 のリングバッファ及び前記第 2 のリングバッファのバッファサイズを変更することを特徴とする請求項 4 に記載のデータ記憶装置。

【請求項 7】 前記バッファ制御手段は、外部から受け付けたアクセス要求の種類及び種類ごとの頻度に基づいて前記第 1、第 2 のリングバッファのバッファサイズを変更することを特徴とする請求項 4 に記載のデータ記憶装置。

【請求項 8】 前記バッファ制御手段は、外部から受け付けたアクセス要求にかかる書き込みデータまたは読み出しデータのサイズに基づいて前記第 1、第 2 のリングバッファのバッファサイズを変更することを特徴とする請求項 4 に記載のデータ記憶装置。

【請求項 9】 所定の記録媒体に対してデータの読み書きを行う際に、当該記録媒体への処理に先立って当該データを一時的に保持するバッファメモリの管理方法であって、

前記記録媒体に対する外部からのアクセス要求を解析するステップと、

前記解析の結果に基づいて、前記バッファメモリのバッファ領域に構築された書き込み要求用のリングバッファ及び読み出しデータ用のリングバッファのバッファサイズを変更するかどうかを判断するステップと、

前記判断に応じて、前記バッファメモリの前記バッファ領域を 2 つの前記リングバッファに分ける仕切りの位置を変更することにより、当該 2 つのリングバッファのバッファサイズを変更するステップと

を含むことを特徴とするバッファメモリの管理方法。

【請求項 10】 前記アクセス要求を解析するステップでは、当該アクセス要求の種類及び種類ごとの頻度を解析し、

前記バッファサイズを変更するかどうかを判断するステップでは、外部からの前記アクセス要求において書き込み要求が多い場合に前記書き込み要求用のリングバッファを拡大するようにバッファサイズを変更すると判断し、読み出し要求

が多い場合に前記読み出しデータ用のリングバッファを拡大するようにバッファサイズを変更すると判断することを特徴とする請求項9に記載のバッファメモリの管理方法。

【請求項11】 前記アクセス要求を解析するステップでは、当該アクセス要求にかかる書き込みデータまたは読み出しデータのサイズを解析し、

前記バッファサイズを変更するかどうかを判断するステップでは、外部からの前記アクセス要求において大きいサイズのデータを書き込むアクセス要求が多い場合に前記書き込み要求用のリングバッファを拡大するようにバッファサイズを変更すると判断し、大きいサイズのデータを読み出すアクセス要求が多い場合に前記読み出しデータ用のリングバッファを拡大するようにバッファサイズを変更すると判断することを特徴とする請求項9に記載のバッファメモリの管理方法。

【請求項12】 前記バッファサイズを変更するステップでは、前記書き込み要求用のリングバッファまたは前記読み出しデータ用のリングバッファのボトムページの位置を変更することにより、2つの前記リングバッファのバッファサイズを変更することを特徴とする請求項9に記載のバッファメモリの管理方法。

【請求項13】 磁気ディスクを記録媒体として用い、マルチメディアコンテンツの記録・再生を行うハードディスク・レコーダにおいて、

前記磁気ディスクに対する書き込み要求を当該磁気ディスクへの処理に先立って一時的に保持する第1のリングバッファ及び前記磁気ディスクに対する読み出し要求に応じて当該磁気ディスクから読み出されたデータを保持する第2のリングバッファを持つバッファ手段と、

マルチメディアコンテンツを記録する動作モードで動作している場合に前記第1のリングバッファに大きなバッファサイズを与え、マルチメディアコンテンツを再生する動作モードで動作している場合に前記第2のリングバッファに大きなバッファサイズを与えるように、前記バッファ手段の記憶領域を動的に割り振るバッファ制御手段と

を備えることを特徴とするハードディスク・レコーダ。

【発明の詳細な説明】

【0001】

**【発明の属する技術分野】**

本発明は、データ記憶装置に関し、特に磁気ディスク等の記録媒体に対してデータの読み書きを行う際に一時的に当該データを保持するバッファメモリの制御に関するものである。

**【0 0 0 2】****【従来の技術】**

磁気ディスク装置（ハードディスク・ドライブ）をはじめとするコンピュータの外部記憶装置では、データの読み書き処理のパフォーマンスを向上する手段として、セクターバッファと呼ばれるバッファメモリ（D R A M : Dynamic Random Access Memory）を用意し、ホストシステムからのデータの書き込み要求（書き込みデータを含む）や1度読み出したデータをキャッシュ（保持）することが常套的に行われている。

一般に、セクターバッファでは、リングバッファと呼ばれるデータ構造が用いられる。リングバッファとは、バッファの始端と終端を連結することで有限長のバッファを無限長として扱う手法である（例えば、特許文献1、2参照）。

**【0 0 0 3】**

図5は、リングバッファの構成を示す概念図である。

図5に示すように、リングバッファでは、セクターバッファの記憶領域をバッファボトムからバッファトップへ順番に使用しながら、連続的なデータとして意味のある固まりをセグメントとして管理する。そして、バッファトップまで使い切った後には、バッファボトムに戻って当該セクターバッファの使用を継続する。これにより意味的（論理的）にリングを形成している。

**【0 0 0 4】**

上述したリングバッファ方式は、セクターバッファを順番に使っていくため、コマンドを順番に処理していくコマンド処理形式には適している。これに対し今日では、データの読み書き処理のパフォーマンスをさらに向上させるために、コマンドのリオーダーリング（Reordering）などが行われるようになった。リオーダーリングとは、データの書き込みまたは読み出しを実行するまでに必要な待ち時間が最短になるように、データの読み出し及び書き込みのコマンドの実行順序を並



べ替えることである。リオーダーリング等によってコマンドの処理順を適宜並べ替えて実行するようになると、書き込み要求と読み出しデータとが混在する形でのセクターバッファにおけるセグメント管理が複雑になってしまう。

#### 【0005】

##### 【特許文献1】

特開 2002-259115号公報

##### 【特許文献2】

特開 2002-170319号公報

#### 【0006】

##### 【発明が解決しようとする課題】

上述したように、リングバッファ方式は、コマンドを順番に処理していくコマンド処理形式には適しているが、コマンドの処理順を適宜並べ替えて実行する場合には、書き込み要求と読み出しデータとが混在する形でのセクターバッファにおけるセグメント管理が複雑になる。

この問題を解消するため、セクターバッファにおいてデータの書き込み要求と読み出しデータとを別個に扱う方式が考えられる。すなわち、セクターバッファ内に書き込み要求をキャッシュするリングバッファと読み出しデータをキャッシュするリングバッファとを設定し、この2つのリングバッファにおけるそれぞれのセグメントを個別に管理する方式である（以下、この方式をデュアル・リングバッファ（Dual Ring Buffer）方式と称す）。書き込み要求と読み出しデータとを個別に管理することにより、データの読み書きに関するコマンドが適宜並べ替えて実行される場合にも、セクターバッファにおけるセグメント管理が複雑になることを回避することができる。

#### 【0007】

しかしこの場合、セクターバッファ内に2つのリングバッファを設定するため、書き込み要求及び読み出しデータのキャッシュに用いられる個々のバッファサイズが小さくなってしまう。すなわち、通常のリングバッファ（以下、シングル・リングバッファ方式と称す）では、キャッシュする内容が制限されないため、外部記憶装置へデータを書き込むアクセスが多ければ書き込み要求が多くキャッ

シュされ、データを読み出すアクセスが多ければ読み出しデータが多くキャッシュされることとなる。言い換えれば、アクセスがデータ書き込みまたはデータ読み出しに偏っている場合には、最大でセクターバッファの記憶容量分の書き込み要求または読み出しデータをキャッシュできることとなる。これに対し、デュアル・リングバッファ方式では、書き込み要求用のリングバッファと読み出しデータ用のリングバッファとが分けられているため、それぞれのリングバッファに割り当てられたバッファサイズ分しかキャッシュすることができない。

#### 【0008】

一般に、キャッシュシステムでは、キャッシュメモリ（バッファメモリ）の記憶容量が大きい方がパフォーマンス向上に寄与することができる。したがって、データの書き込みまたは読み出しに偏ったアクセスがある場合、セクターバッファにおいてデュアル・リングバッファ方式を採る外部記憶装置は、シングル・リングバッファ方式を採る外部記憶装置よりもパフォーマンスが悪化してしまう。

コンピュータで扱うデータがマルチメディア化し、また動画や音声を磁気ディスクに記録するハードディスク・レコーダが普及しつつある今日、画像データや音声データといったサイズの大きいデータを1度に書き込みあるいは読み出すような偏ったアクセスが行われる機会も増加している。

#### 【0009】

そこで本発明は、バッファメモリのデータ構造としてデュアル・リングバッファ方式を採用しながら、データの書き込みまたは読み出しに偏ったアクセスが行われる場合にも、記憶装置におけるパフォーマンスの悪化を回避することを目的とする。

#### 【0010】

##### 【課題を解決するための手段】

上記の目的を達成するため、本発明は、次のように構成されたデータ記憶装置として実現される。すなわち、このデータ記憶装置は、データを記録する記録媒体と、この記録媒体に対する外部からの書き込み要求を記録媒体への処理に先立って一時的に保持する第1のリングバッファ及び記録媒体に対する外部からの読み出し要求に応じて記録媒体から読み出されたデータを保持する第2のリングバ

ッファを持つバッファ手段と、記録媒体への外部からのアクセスパターンに応じて、このバッファ手段における第1のリングバッファ及び第2のリングバッファのバッファサイズを変更するバッファ制御手段とを備えることを特徴とする。

#### 【0011】

より好ましくは、このバッファ手段は、バッファ領域を第1のリングバッファと第2のリングバッファとに分割したRAM (Random Access Memory) で構成される。そして、第1のリングバッファのボトムページまたは第2のリングバッファのボトムページの位置を変更することにより、各リングバッファのバッファサイズが変更される。

また、バッファサイズの変更は、外部から受け付けたアクセス要求の種類及び種類ごとの頻度に基づいて、もしくは、外部から受け付けたアクセス要求にかかる書き込みデータまたは読み出しデータのサイズに基づいて行うことができる。

#### 【0012】

また、本発明は、記録媒体へのデータの読み書きに先立って当該データを一時的に保持するバッファメモリの管理方法としても実現される。このバッファメモリの管理方法は、記録媒体に対する外部からのアクセス要求を解析するステップと、この解析の結果に基づいて、バッファメモリのバッファ領域に構築された書き込み要求用のリングバッファ及び読み出しデータ用のリングバッファのバッファサイズを変更するかどうかを判断するステップと、この判断に応じて、バッファメモリのバッファ領域を2つのリングバッファに分ける仕切りの位置を変更することにより、これら2つのリングバッファのバッファサイズを変更するステップとを含むことを特徴とする。

#### 【0013】

より詳細には、外部からの前記アクセス要求において書き込み要求が多い場合に書き込み要求用のリングバッファを拡大するようにバッファサイズを変更し、読み出し要求が多い場合に読み出しデータ用のリングバッファを拡大するようにバッファサイズを変更する。あるいは、外部からの前記アクセス要求において大きいサイズのデータを書き込むアクセス要求が多い場合に書き込み要求用のリングバッファを拡大するようにバッファサイズを変更し、大きいサイズのデータを

読み出すアクセス要求が多い場合に読み出しデータ用のリングバッファを拡大するようにバッファサイズを変更する。

#### 【0014】

さらにまた、上記の目的を達成する本発明は、次のように構成された磁気ディスク装置として実現される。すなわち、この磁気ディスク装置は、磁氣的にデータを記録する磁気ディスクと、この磁気ディスクに対して読み書きされるデータを磁気ディスクへの処理に先立って一時的に保持するバッファメモリとを備え、このバッファメモリは、記憶領域を、外部から受け付けた磁気ディスクへの書き込み要求を格納する第1のリングバッファと、磁気ディスクから読み出したデータを格納する第2のリングバッファとに分け、第1のリングバッファ及び第2のリングバッファを分ける仕切りの位置を可変とすることにより、書き込み用領域及び読み出し用領域のバッファサイズを可変としたことを特徴とする。

#### 【0015】

さらに本発明は、磁気ディスクを記録媒体として用い、マルチメディアコンテンツの記録・再生を行う、次のように構成されたハードディスク・レコーダとしても実現される。このハードディスク・レコーダは、磁気ディスクに対する書き込み要求をこの磁気ディスクへの処理に先立って一時的に保持する第1のリングバッファ及び磁気ディスクに対する読み出し要求に応じてこの磁気ディスクから読み出されたデータを保持する第2のリングバッファを持つバッファ手段と、マルチメディアコンテンツを記録する動作モードで動作している場合に第1のリングバッファに大きなバッファサイズを与え、マルチメディアコンテンツを再生する動作モードで動作している場合に第2のリングバッファに大きなバッファサイズを与えるように、バッファ手段の記憶領域を動的に割り振るバッファ制御手段とを備えることを特徴とする。

#### 【0016】

##### 【発明の実施の形態】

以下、添付図面に示す実施の形態に基づいて、この発明を詳細に説明する。本実施の形態では、外部記憶装置として磁気ディスクを記録媒体に用いるハードディスク装置を例にして説明する。

図1は、ハードディスク装置100の主要部の構成を示すブロック図である。

ハードディスク装置100は、スピンドルモータ102によって回転駆動される磁気ディスク101上を磁気ヘッド103がシークしかつ所定のトラック（位置）に留まって磁気ディスク101に対してデータを書き込み、または磁気ディスク101に書き込まれたデータを読み出す外部記憶装置である。磁気ディスク101は、データを磁気情報として記録する記録媒体であり、必要に応じて単数または複数搭載されるが、図1においては、単数の例を示している。

#### 【0017】

磁気ディスク101は、ハードディスク装置100が動作しているとき、スピンドルモータ102のスピンドル軸を中心にして回転駆動され、ハードディスク装置100が非動作のとき、回転停止（静止）する。磁気ヘッド103は、アクチュエータ104の先端部に磁気ディスク101の表裏面に対応して2つ保持されており、磁気ディスク101に対してデータの書き込み及び読み出しを実行する。アクチュエータ104は、デジタル／アナログ変換器（DAC）106及びボイス・コイル・モータ（VCM）ドライバ107を介してMPU（Micro Processing Unit）109に制御されたボイス・コイル・モータ（VCM）105によって駆動される。

#### 【0018】

リード・ライト・チャネル108は、データの読み書き処理を実行する。つまり、データの書き込みでは、HDC（ハードディスク・コントローラ）110を介してホストシステム（コンピュータ）から転送された書き込みデータを書き込み信号（電流）に変換して磁気ヘッド103に供給する。磁気ヘッド103は、この書き込み電流に基づいて磁気ディスク101に対してデータの書き込みを実行する。一方、データの読み出しでは、磁気ディスク101から読み取った読み取り信号（電流）をデジタル・データに変換して、HDC110を介してホストシステムに出力する。

#### 【0019】

HDC110は、ハードディスク装置100のインターフェイスとしての機能を有している。その機能の1つとして、ホストシステムから転送された書き込み

データを受けると共にリード・ライト・チャンネル108に転送する。また、リード・ライト・チャンネル108から転送される読み出しデータをホストシステムに転送する。さらに、ホストシステムからの指示コマンド等を受けてMPU109に転送する。

#### 【0020】

MPU109は、ハードディスク装置100の制御を担う。

DRAM111は、MPU109がハードディスク装置100を制御するための制御プログラムを格納すると共に、データの読み書きにおけるセクターバッファ（バッファメモリ）として用いられる。したがって、HDC110は、ホストシステムから転送された書き込み要求（書き込みデータを含む）を、一旦DRAM111にキャッシュし、この時点で、ホストシステムに対する当該書き込み要求に対する完了通知（コマンド・コンプリート）を発行、通知する。その後、適当なタイミングで、キャッシュされている書き込みデータをリード・ライト・チャンネル108に転送し、書き込み要求を実行する。また、HDC110は、リード・ライト・チャンネル108から転送された読み出しデータをDRAM111にキャッシュしておき、ホストシステムから同一データの読み出し要求を受け付けた場合に、DRAM111にキャッシュされている当該データをホストシステムへ返す。

#### 【0021】

図2は、DRAM111に設定されるセクターバッファの構成例を示す図である。

図2に示すように、本実施の形態によるセクターバッファ10は、データ構造としてデュアル・リングバッファ方式を採用しており、セクターバッファ内部を読み取りデータ用の領域であるリングバッファ（以下、BUF\_R）11と、書き込み要求用の領域であるリングバッファ（以下、BUF\_W）12とに分けている。図示の例では、BUF\_W12のボトムページ（以下、BPAGE）13をBUF\_R11とBUF\_W12との仕切りとして用いている。すなわち、セクターバッファのバッファボトムからBPAGE13までをBUF\_R11として用い、セクターバッファのバッファトップからBPAGE13までをBUF\_

W12として用いている。なお、図示の例では、BUF\_W12のボトムページを仕切り(BPAGE13)としているが、BUF\_R11がバッファトップ側にあり、BUF\_W12がバッファボトム側にある場合は、BUF\_R11のボトムページを仕切り(BPAGE13)として用いることは言うまでもない。

#### 【0022】

また、本実施の形態は、MPU109の制御により、BPAGE13の位置を変更可能としている。すなわち、本実施の形態においてMPU109は、ホストシステムからハードディスク装置100(磁気ディスク101)へのアクセスパターン(書き込み要求と読み出し要求の頻度や扱うデータのサイズ等のパターン)に応じて、セクターバッファ10全体の記憶領域を、BUF\_R11及びBUF\_W12に、動的かつ適切に割り振る制御手段である。

#### 【0023】

図3は、セクターバッファ10の制御手段としてのMPU109の機能を示すブロック図である。

図3を参照すると、本実施の形態によるMPU109は、ホストシステムから来たアクセス要求(以下、アクセスコマンド)を受け付ける割り込み検知部(Interrupt Handler)20と、セクターバッファ10のセグメントを管理し制御するセグメント制御部(Segment Control Routine)30と、セクターバッファ10の構成を変更するためのコマンドを発行するコマンド実行制御部(Command Dispatcher)40と、イベント検知部(Event Handler)50とを備える。また、アクセスコマンドを解析してセクターバッファ10のBPAGE13を移動するかどうかを判断するためのコマンド解析部60と、セクターバッファ10を用いた処理にエラーが起きた場合に、これを検知し復帰処理(リトライ)を行うためのエラー復帰部(Error Recovery Handler)70を備える。

これらの機能は、DRAM111に格納されたプログラム(マイクロコード)にてMPU109を制御することにより実現される。このプログラムは、例えば磁気ディスク101に書き込まれており、ハードディスク装置100の起動時にDRAM111に転送される。

#### 【0024】

上記構成において、割り込み検知部 2 0 は、H D C 1 1 0 から転送されたホストシステムからのアクセスコマンドを受け付け、要求の種類（データの書き込みまたは読み出し）に応じた処理を行う。書き込み要求を受け付けた場合、割り込み検知部 2 0 は、書き込み要求にかかる書き込みデータをセクターバッファ 1 0 に格納し、セグメント制御部 3 0 を呼び出して処理を開始させる。この際、さらに次の書き込み要求が来ても受け入れられる余裕がセクターバッファ 1 0 にある場合、セクターバッファ 1 0 から割り込み検知部 2 0 へ書き込み転送完了通知（Write Transfer Complete）が発行される。この通知を受けた割り込み検知部 2 0 は、ホストシステムに対して完了通知（Command Complete）を返送する。また、割り込み検知部 2 0 は、データの読み出し要求を受け付けた場合、セグメント制御部 3 0 を呼び出して処理を開始させる。

#### 【 0 0 2 5 】

セグメント制御部 3 0 は、セクターバッファ 1 0 に格納されているデータと割り込み検知部 2 0 が受け付けたアクセスコマンドとの関連を調べ、当該要求に応じた処理を行う。割り込み検知部 2 0 の受け付けたアクセスコマンドが書き込み要求である場合、例えば、当該書き込み要求にかかる書き込みデータの書き込み先と同じアドレスのデータが B U F \_ R 1 1 内に保持されているならば、このデータを破棄する処理を行う。また、割り込み検知部 2 0 の受け付けたアクセスコマンドが読み出し要求である場合、当該読み出し要求にかかるデータが B U F \_ R 1 1 あるいは B U F \_ W 1 2 内に保持されている（ヒットした）ならば、当該データを読み出してホストシステムに送信する。当該読み出し要求にかかるデータが B U F \_ R 1 1 及び B U F \_ W 1 2 内に保持されていないならば、磁気ディスク 1 0 1 からの当該データの読み出しが必要となるので、コマンド実行制御部 4 0 を呼び出す。

さらに、本実施の形態のセグメント制御部 3 0 は、所定の場合にセクターバッファ 1 0 における B P A G E 1 3 を移動する。この処理の詳細については後述する。

#### 【 0 0 2 6 】

コマンド実行制御部 4 0 は、磁気ヘッド 1 0 3 のシークや管理テーブルの生成



、データの読み出し、書き込み等の指示を、適切なタイミングで発行する。磁気ディスク 101 からデータを読み出すためにセグメント制御部 30 から呼び出された場合は、直ちに当該データの読み出しの指示を発行する。

さらに、本実施の形態のコマンド実行制御部 40 は、所定の場合にセクターバッファ 10 における B P A G E 1 3 を移動するために必要な処理を行う。この処理の詳細については後述する。

#### 【0027】

イベント検知部 50 は、コマンド実行制御部 40 から発行された各種の指示をイベントとして検出し、当該指示を実現して磁気ディスク 101 に対してデータの読み書きを行うための適切なルーチンを起動する。これにより、実際にデータが磁気ディスク 101 に書き込まれ、あるいは磁気ディスク 101 からデータの読み出しが行われる。

磁気ディスク 101 から読み出されたデータは、セクターバッファ 10 の B U F \_ R 1 1 に格納されると共に、ホストシステムに対して送信される。

#### 【0028】

コマンド解析部 60 は、割り込み検知部 20 が受け付けたアクセスコマンドを解析して、B P A G E 1 3 を移動するかどうかを判断する。コマンド解析及び判断の方法としては、受け付けたアクセスコマンドの種類や頻度に基づいた次のような方法を採用することができる。

(1) アクセスコマンドの種類（読み出し要求か書き込み要求か）及びコマンド長（読み出しまたは書き込まれるデータのサイズ）を解析する。そして、一定以上のコマンド長の読み出し要求が一定個数以上来た場合（具体的には、64セクタを越えるサイズのデータの読み出し要求が3つ以上来た場合等）、B U F \_ R 1 1 を拡張するように B P A G E 1 3 を移動する（図2の例では B P A G E 1 3 をバッファトップ側へ移動する）と判断する。また、書き込み要求が来た場合に B U F \_ W 1 2 を拡張するように（元へ戻すように） B P A G E 1 3 を移動すると判断する。

#### 【0029】

(2) アクセスコマンドの種類（読み出し要求か書き込み要求か）を解析する。

そして、書き込み要求をキャッシュしていった BUF\_W 1 2 の空き領域が少なくなった場合に BUF\_W 1 2 を拡張すると判断する（BUF\_W 1 2 の空き領域の情報はセグメント制御部 3 0 から得ることができる）。なお、この場合、初期的には BUF\_W 1 2 のサイズを小さめに設定しておく。

#### 【 0 0 3 0 】

(3) アクセスコマンドの種類（読み出し要求か書き込み要求か）を解析し、読み出し要求と書き込み要求とに分けて受け付けた個数をカウントする。そして、カウント値に一定以上の差が生じたならば、その差に応じて（読み出し要求が多ければ BUF\_R 1 1 を拡張し、書き込み要求が多ければ BUF\_W 1 2 を拡張するように）B PAGE 1 3 を移動すると判断する。この場合、どれだけの差が生じた場合に B PAGE 1 3 をどちらに移動するかを決定する閾値を予め設定し、例えば DRAM 1 1 1 に保持しておく。

この方法の変形として、さらにコマンド長（読み出しまたは書き込まれるデータのサイズ）をも解析し、読み出し要求及び書き込み要求の受け付け個数とそれぞれのコマンド長とを複合的に判断して B PAGE 1 3 を移動すると判断することもできる。

#### 【 0 0 3 1 】

(4) アクセスコマンドの種類（読み出し要求か書き込み要求か）を解析する。そして、BUF\_R 1 1 及び BUF\_W 1 2 のそれぞれにキャッシュされたデータのヒット率を算出し、より効果的なバッファサイズとなるように（ヒット率の高い方に大きいサイズを割り振るように）、B PAGE 1 3 を移動すると判断する。

#### 【 0 0 3 2 】

これらの他、ハードディスク装置 1 0 0 が、コンピュータの外部記憶装置としてのハードディスク・ドライブではなく、動画データや音声データ等のマルチメディアコンテンツを記録・再生するためのハードディスク・レコーダである場合、マルチメディアコンテンツを記録する際と再生する際とでは、磁気ディスクへのアクセスパターンが大きく異なる。すなわち、記録する際は大きなデータの書き込み要求が多く、再生する際は大きなデータの読み出し要求が多い。そこで、

ハードディスク装置 100 の動作モードを監視し、記録動作の際には BUF\_W12 が大きくなるように BPAGE 13 を移動し、再生動作の際には BUF\_R11 が大きくなるように BPAGE 13 を移動するように制御することもできる。

#### 【0033】

以上のような方法で、コマンド解析部 60 が BPAGE 13 を移動するかどうかを判断した後、コマンド実行制御部 40 から発行されるコマンドによってセクターバッファ 10 の状態が BPAGE 13 を移動可能な状態かどうか調査され、移動可能であれば、セグメント制御部 30 により BPAGE 13 の位置が変更される。ここで、BPAGE 13 の移動については、BUF\_R11 と BUF\_W12 とが何段階かの所定の割合となるように BPAGE 13 の位置をいくつか予め設定しておき、段階的に移動させることが可能である。

#### 【0034】

エラー復帰部 70 は、上記割り込み検知部 20 乃至イベント検知部 50 による一連の読み書き処理とは別に、当該読み書き処理においてエラーが発生した場合に起動し、予め用意されたエラー復帰手順 (Error Recovery Procedure) に従ってデータの書き込み処理あるいは読み出し処理を再施行する。

#### 【0035】

次に、ハードディスク装置 100 がホストシステムからアクセスコマンドを受け取った際のセクターバッファ 10 に対する制御の動作を説明する。

図 4 は、MPU 109 による処理の流れを説明するフローチャートである。

図 4 に示すように、ハードディスク装置 100 がホストシステムからのアクセスコマンドを受け取り、当該アクセスコマンドが HDC 110 から MPU 109 へ転送されると (ステップ 401)、まずコマンド解析部 60 により当該アクセスコマンドが解析される。そして、解析結果に基づいてセクターバッファ 10 の BPAGE 13 を移動するかどうかの判断がなされる (ステップ 402)。ここで、BPAGE 13 を移動すると判断された場合、マイクロコードが参照するフラグである BPAGE change request がオンとなる。

#### 【0036】

次に、セグメント制御部 30 により、当該アクセスコマンドの BPAGE change request の状態がチェックされる。BPAGE change request がオフであれば、通常通り当該アクセスコマンドが実行されて処理が終了する（ステップ 403、404）。

一方、BPAGE change request がオンであれば、次に、実際に BPAGE 13 の書き換えが行われるが、この際、処理中のアクセスコマンドとセクターバッファ 10 の状態とに配慮する必要がある。すなわち、BPAGE 13 が BUF\_W12 中のセグメントに含まれていないならば BPAGE 13 の書き換えを行うことができるが、BPAGE 13 が BUF\_W12 中のセグメントに含まれている場合は、BPAGE 13 を書き換えることによって BUF\_W12 の当該セグメントが影響を受けるため、BPAGE 13 の書き換えを行うことはできない。

#### 【0037】

具体的な動作としては、コマンド実行制御部 40 により、まず num\_pend wrt という変数を調べるコマンドが発行され、このコマンドが所定のルーチンにて実行される。そして、num\_pend wrt の値に応じて、次にセグメント制御部 30 により、BPAGE 13 が所定の書き込み要求のセグメントに含まれているかどうか調べられる（ステップ 405）。ここで、num\_pend wrt とは、セクターバッファ 10 に保持されていながら、未だ磁気ディスク 101 への書き込みが終了していない書き込み要求の数を示す変数である。

#### 【0038】

num\_pend wrt の値が 0 であれば、未だ磁気ディスク 101 への書き込みが終了していない書き込み要求のセグメントは存在しないので、BPAGE 13 を含む BUF\_W12 中のセグメントを検索する必要はない。一方、num\_pend wrt が 0 でない場合は、BPAGE 13 を含むセグメントを検索する。

BPAGE 13 を含むセグメントを検索した結果、そのようなセグメントが存在するならば、BPAGE 13 を書き換えることはできないので、通常通りアクセスコマンドが実行されて処理が終了する（ステップ 404）。

#### 【0039】

BPAGE 13 を含むセグメントが存在しない場合は、BPAGE 13 の書き

換えが可能であるので、まずコマンド実行制御部40からページ・ディスカード (PAGE Discard) を行うコマンドが発行され、所定のルーチンによりページ・ディスカードが実行される。(ステップ406)。ページ・ディスカードとは、キャッシュが不要なセグメントを該当ページから破棄する処理である。例えばBUF\_\_W12に保持されている書き込み要求の内、既に磁気ディスク101への書き込みが終わったもの等である。

そして、セグメント制御部30により、BPAGE13が書き換えられて位置が変更され、これに伴ってBUF\_\_R11及びBUF\_\_W12の管理パラメータが変更される(ステップ407)。その後、通常通りアクセスコマンドが実行されて処理が終了する(ステップ404)。

#### 【0040】

以上の動作は、ハードディスク装置100がホストシステムからアクセスコマンドを受け取る度に実行される。ステップ405の判断に示したように、本実施の形態では、コマンド解析の結果からBPAGE13を移動してBUF\_\_R11及びBUF\_\_W12のバッファサイズを変更した方が好ましい場合であっても、セクターバッファ10の状態によっては、すなわち、BPAGE13を含む書き込み要求のセグメントが存在する場合は、BPAGE13の書き換えを行わず、次の機会まで(別のアクセスコマンドを受け付けてステップ401以降の動作サイクルを開始するまで)待つ。これは、次のアクセスコマンドを受け付けるまでの間に、ハードディスク装置100におけるバックグラウンドでの通常の処理として、BUF\_\_W12に保持されている書き込み要求が実行され、BPAGE13を含む書き込み要求のセグメントが無くなれば、BPAGE13の書き換えが可能となるためである。

#### 【0041】

上述したように、本実施の形態では、セクターバッファ10の構成を、読み出しデータをキャッシュするリングバッファであるBUF\_\_R11とホストシステムからの書き込み要求をキャッシュするリングバッファであるBUF\_\_W12とに分割したデュアル・リングバッファ方式とする。そして、ホストシステムからハードディスク装置100(磁気ディスク101)へのアクセスパターンに応じ

て、書き込み要求が多い場合やそのデータサイズが大きい場合に BUF\_W12 を拡張するように、読み出し要求が多い場合やそのデータサイズが大きい場合に BUF\_R11 を拡張するように、各々のバッファサイズを変更する。セクターバッファ 10 全体のバッファサイズは決まっているので、BUF\_R11 と BUF\_W12 とを分ける仕切りである、BUF\_R11 または BUF\_W12 のボトムページ (B PAGE 13) の位置を変更することにより、BUF\_R11 及び BUF\_W12 のバッファサイズを変更している。

#### 【0042】

これにより、読み出しデータに対するキャッシュ容量を多く必要な場合には BUF\_R11 のバッファサイズを大きくし、書き込み要求に対するキャッシュ容量を多く必要な場合には BUF\_W12 のバッファサイズを大きくすることができるので、セクターバッファ 10 のデータ構造としてデュアル・リングバッファ方式を用いていても、書き込み要求及び読み出しデータのキャッシュに用いられる個々のバッファサイズがシングル・リングバッファ方式の場合と比べて小さくなってしまおうという影響を低減することができる。

#### 【0043】

##### 【発明の効果】

以上説明したように、本発明によれば、バッファメモリのデータ構造としてデュアル・リングバッファ方式を採用しながら、データの書き込みまたは読み出しに偏ったアクセスが行われる場合にも、記憶装置におけるパフォーマンスの悪化を回避することが可能となる。

##### 【図面の簡単な説明】

【図 1】 本実施の形態によるハードディスク装置の主要構成を示すブロック図である。

【図 2】 本実施の形態における DRAM に設定されるセクターバッファの構成例を示す図である。

【図 3】 本実施の形態におけるセクターバッファの制御手段としての MPU の機能を示すブロック図である。

【図 4】 本実施の形態における MPU による処理の流れを説明するフロー

チャートである。

【図5】 リングバッファの構成を示す概念図である。

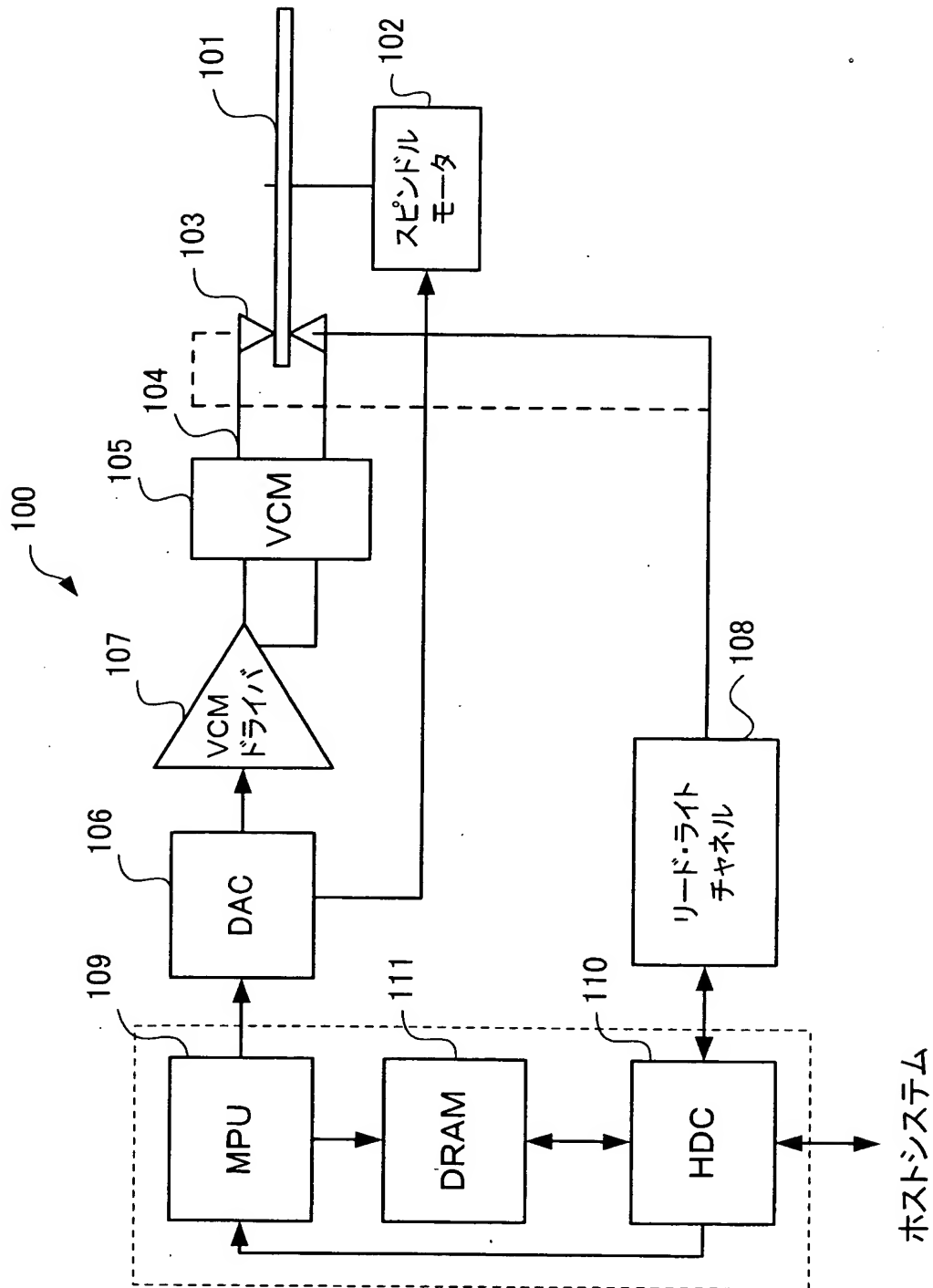
【符号の説明】

10…セクターバッファ、11…BUF\_\_R（リングバッファ）、12…BUF\_\_W（リングバッファ）、13…BPAGE、20…割り込み検知部（Interrupt Handler）、30…セグメント制御部（Segment Control Routine）、40…コマンド実行制御部（Command Dispatcher）、50…イベント検知部（Event Handler）、60…コマンド解析部、70…エラー復帰部（Error Recovery Handler）、100…ハードディスク装置、101…磁気ディスク、102…スピンドルモータ、103…磁気ヘッド、104…アクチュエータ、105…ボイス・コイル・モータ（VCM）、106…デジタル／アナログ変換器（DAC）、107…ボイス・コイル・モータドライバ、108…リード・ライト・チャネル、109…MPU、110…HDC（ハードディスク・コントローラ）、111…DRAM

【書類名】

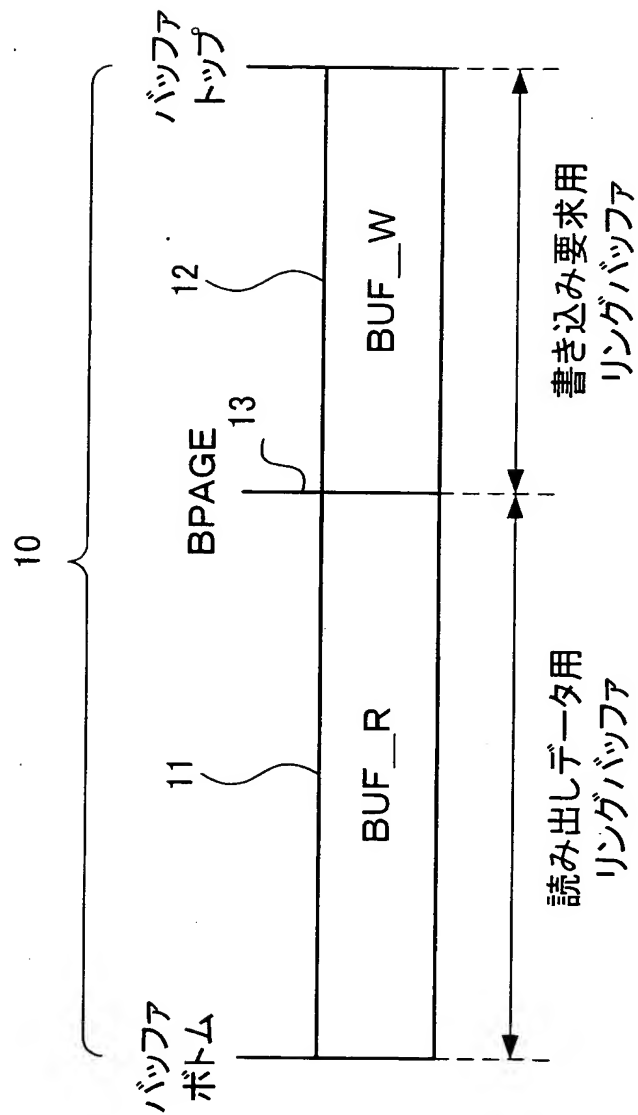
図面

【図 1】

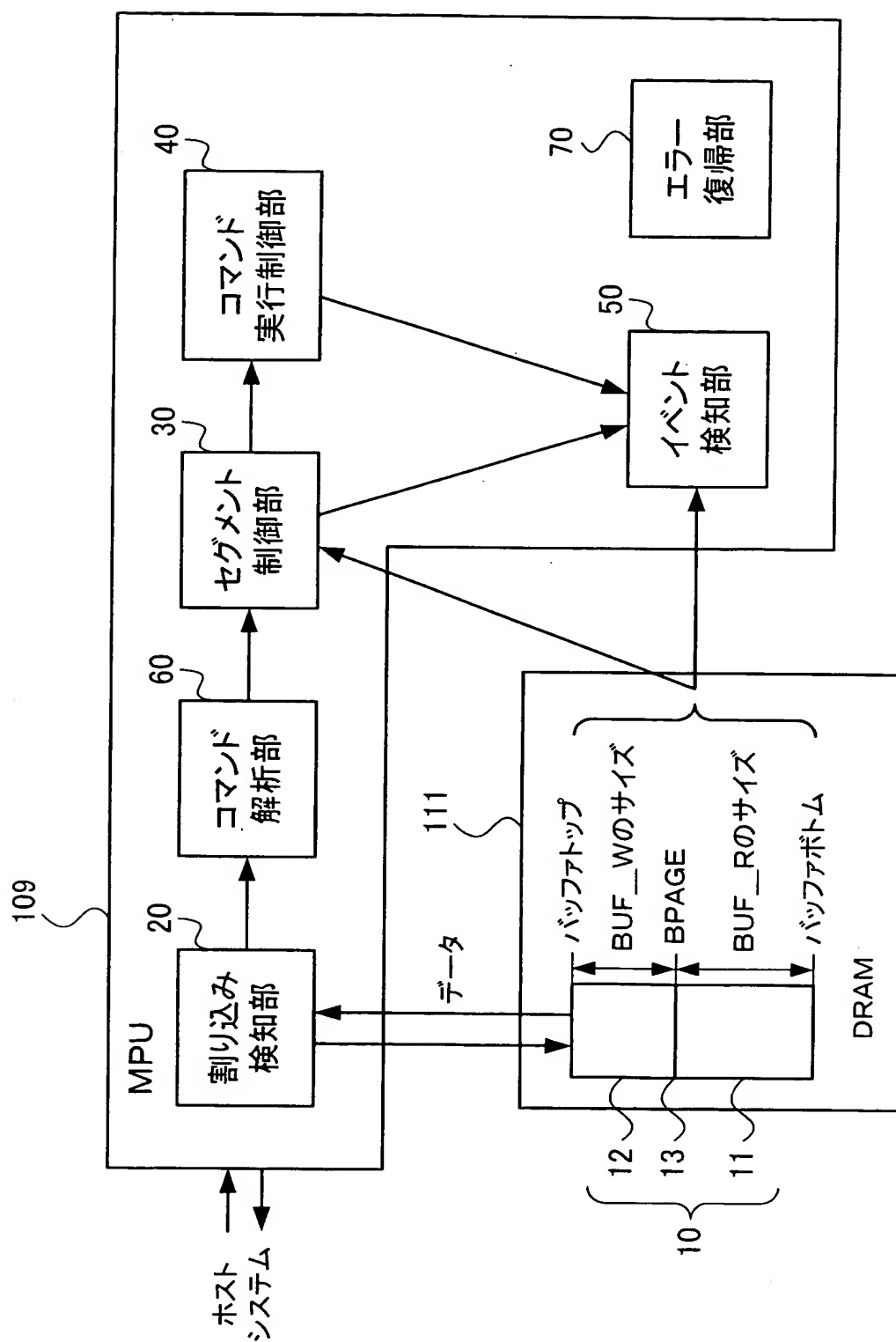




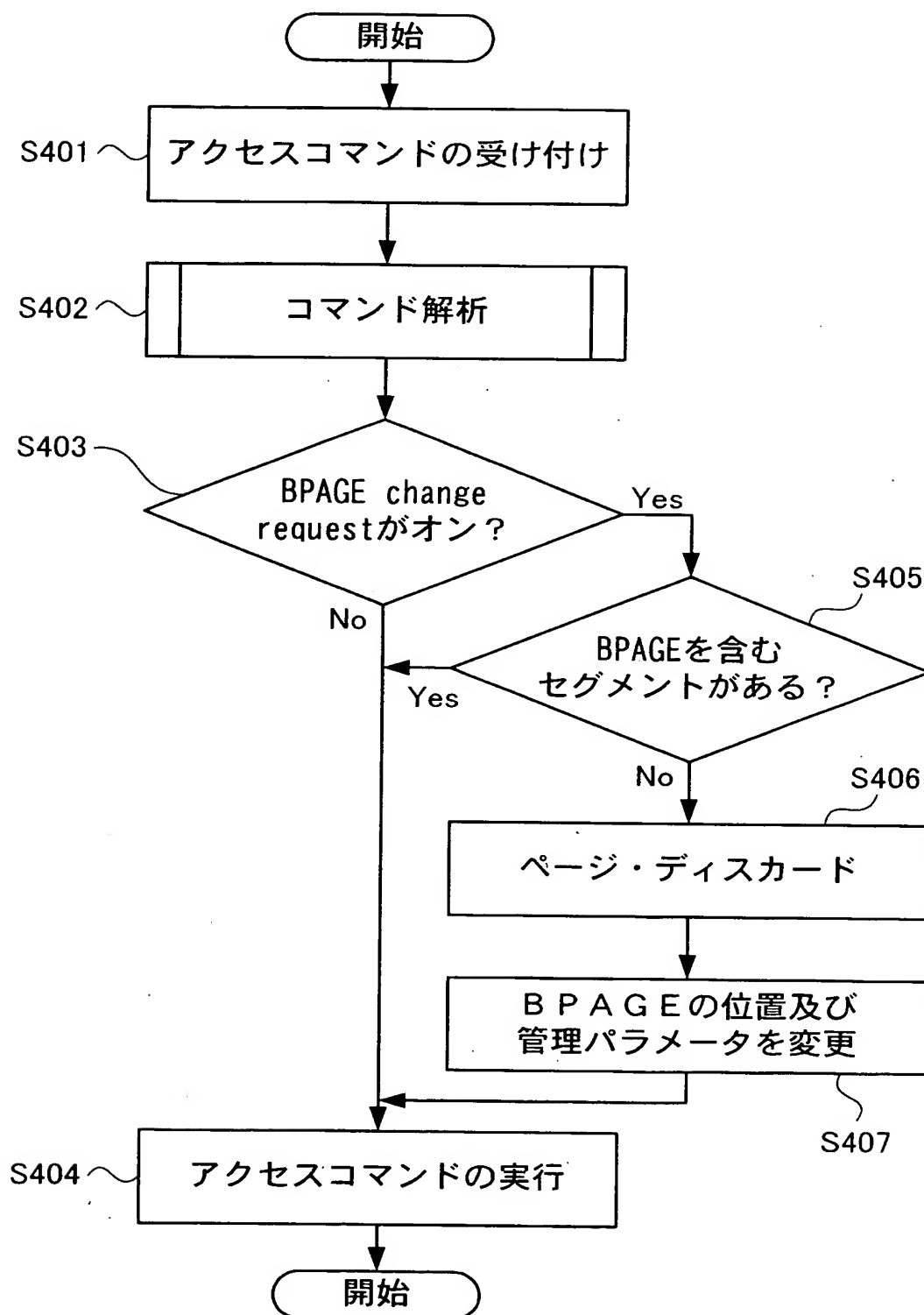
【図 2】



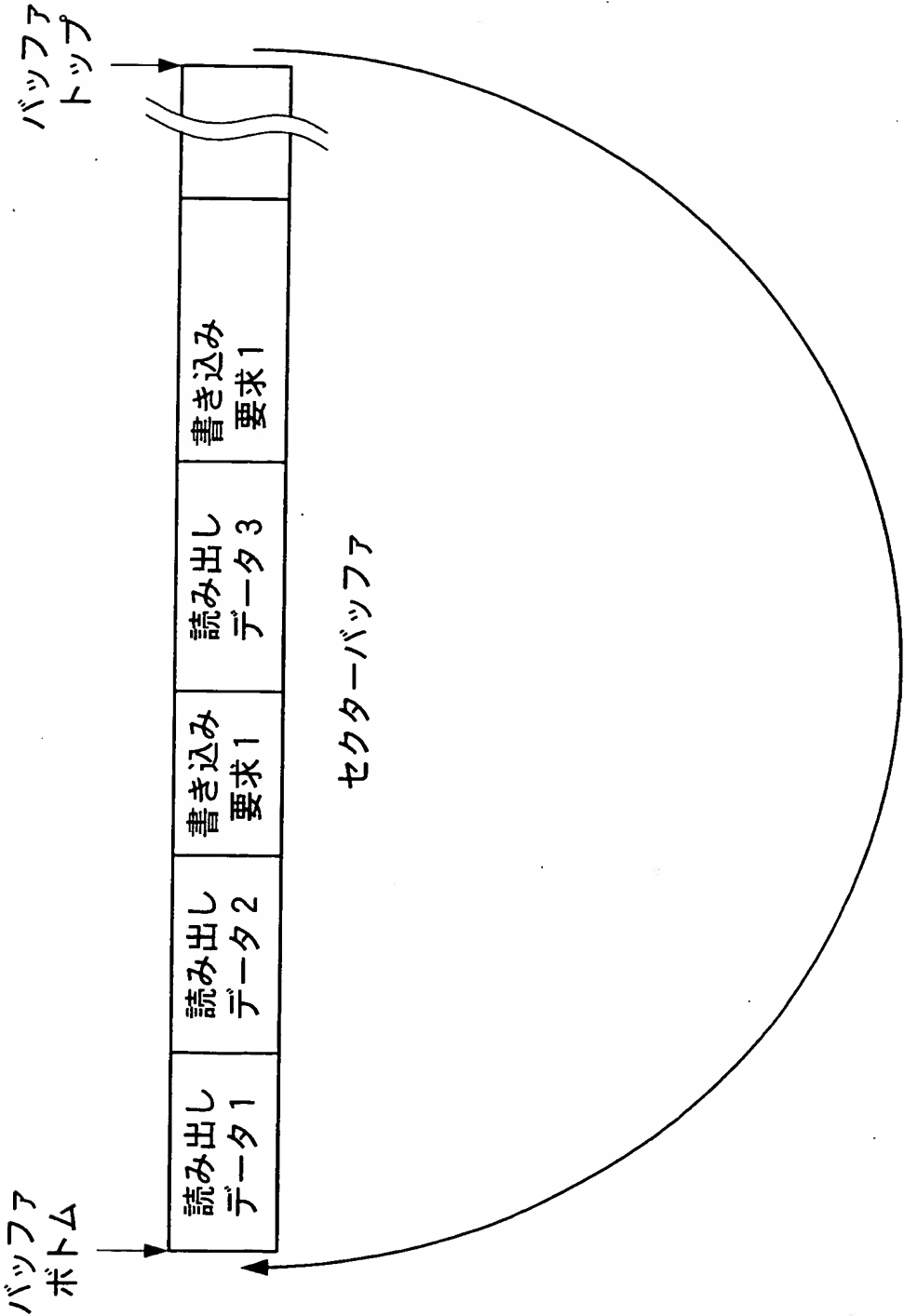
【図 3】



【図 4】



【図 5】



【書類名】 要約書

【要約】

【課題】 データ構造としてデュアル・リングバッファ方式を採るバッファメモリにおいて、データの書き込みまたは読み出しに偏ったアクセスが行われる場合にも、記憶装置におけるパフォーマンスの悪化を回避する。

【解決手段】 記憶装置のDRAM111にアクセス要求をキャッシュするセクターバッファ10を設ける。セクターバッファ10は、記録媒体に対する外部からの書き込み要求を記録媒体への処理に先立って一時的に保持する第1のリングバッファ(BUF\_\_W)12及び記録媒体に対する外部からの読み出し要求に応じてこの記録媒体から読み出されたデータを保持する第2のリングバッファ(BUF\_\_R)11を持つ。記憶装置のMPU109は、記録媒体への外部からのアクセスパターンに応じてBUF\_\_R11またはBUF\_\_W12のBPAGE13を移動することにより、BUF\_\_R11及びBUF\_\_W12のバッファサイズを変更する。

【選択図】 図2

## 認定・付加情報

特許出願の番号	特願 2002-338497
受付番号	50201762356
書類名	特許願
担当官	塩野 実 2151
作成日	平成14年11月25日

## &lt;認定情報・付加情報&gt;

## 【特許出願人】

【識別番号】	390009531
【住所又は居所】	アメリカ合衆国10504、ニューヨーク州 アーモンク ニュー オーチャード ロード
【氏名又は名称】	インターナショナル・ビジネス・マシーンズ・コーポレーション

## 【代理人】

【識別番号】	100086243
【住所又は居所】	神奈川県大和市下鶴間1623番地14 日本アイ・ビー・エム株式会社 大和事業所内
【氏名又は名称】	坂口 博

## 【代理人】

【識別番号】	100091568
【住所又は居所】	神奈川県大和市下鶴間1623番地14 日本アイ・ビー・エム株式会社 大和事業所内
【氏名又は名称】	市位 嘉宏

## 【代理人】

【識別番号】	100108501
【住所又は居所】	神奈川県大和市下鶴間1623番14 日本アイ・ビー・エム株式会社 知的所有権
【氏名又は名称】	上野 剛史

## 【復代理人】

申請人

【識別番号】	100104880
【住所又は居所】	東京都港区赤坂5-4-11 山口建設第2ビル 6F セリオ国際特許事務所
【氏名又は名称】	古部 次郎

## 【選任した復代理人】

【識別番号】	100118201
--------	-----------

次頁有

認定・付加情報 (続き)

【住所又は居所】 東京都港区赤坂 5-4-11 山口建設第二ビル  
6 F セリオ国際特許事務所  
【氏名又は名称】 千田 武

次頁無

特願 2 0 0 2 - 3 3 8 4 9 7

出 願 人 履 歴 情 報

識別番号

[ 3 9 0 0 0 9 5 3 1 ]

1. 変更年月日 2 0 0 0 年 5 月 1 6 日  
[変更理由] 名称変更  
住 所 アメリカ合衆国 1 0 5 0 4 、 ニューヨーク州 アーモンク ( 番地なし)  
氏 名 インターナショナル・ビジネス・マシーンズ・コーポレーション
2. 変更年月日 2 0 0 2 年 6 月 3 日  
[変更理由] 住所変更  
住 所 アメリカ合衆国 1 0 5 0 4 、 ニューヨーク州 アーモンク ニュー オーチャード ロード  
氏 名 インターナショナル・ビジネス・マシーンズ・コーポレーション